PROTEIN ENGINEERING OF A FUNGAL LIPASE

Towards expression of a recombinant Candida rugosa lipase

LILIA ALBERGHINA and MARINA LOTTI
Dipartimento Fisiologia e Biochimica Generali
Sezione Biochimica Comparata
Universita' degli Studi di Milano
Via Celoria 26, 20133 Milano, Italy

1. Introduction

Cloning and overexpression of recombinant proteins in host cells make available large amounts of purified proteins, natural or modified by protein engineering techniques to improve their catalytic performances. This approach is especially suitable to purify isoforms of *Candida rugosa* lipase (CRL) that differ in biochemical and possibly in catalytic properties but are still similar enough to be only laboriously separated, as described in another paper of this book.

We have cloned a large family of lipase-encoding genes that are likely expressed by the yeast in inequal amounts and under different growth conditions [1,2]. CRL isozymes are similar in molecular weight and sequence (about 80% pairwise identity in the amino acid sequences) but differ in pI and glycosylation degree; variability in catalytic activity is suggested by experimental evidence [3]. The product of the *LIP1* gene is likely to be the major enzyme form and its sequence and three-dimensional structure have been elucidated [1,4].

For these reasons we decided to express *LIP1* in *Saccharomyces cerevisiae*. In this attempt we had to face unexpected issues concerning i) the translability/stability of the recombinant protein in the host cells and ii) the non universal genetic code utilized by *Candida* cells. These topics are the subject of this paper.

2. Heterologous expression of LIP1 lipase

S. cerevisiae was the host selected for expression of cloned lipase I. In our laboratory *S. cerevisiae* is currently used for expressing rDNA proteins from bacteria, yeasts, plant and mammals [5] at levels as high as 15% of total cell proteins. In the pharmaceutical industry, this organism is widely employed for the production of proteins like hepatitis B vaccine and human insuline. Like

F. X. Malcata (ed.), Engineering of with Lipases, 219–228.

mammalian cells, yeast is able to carry out several post-translational modifications but it is as easy to handle as are bacterial cells. Information about protein processing in *Candida* is very poor, but it is known for example, that the lipase proteins are N-glycosylated; thus a *Candida* protein should be correctly glycosylated by *Saccharomyces* cells. This point seems to be of importance also for lipase function, since a role of the sugar mojety on the stabilization of the CRL active conformation has been observed [4].

As the expression vector we used the *E. coli-S. cerevisiae* shuttle plasmid pEMBLYex4 [6] that carries the upstream activating sequence UASGal isolated from yeast *GAL1-GAL10* genes. Expression of any gene cloned downstream of UASGal is induced up to 1000 fold by the presence of galactose in the culture medium whereas it is repressed by glucose through a mechanism known as catabolite repression [7].

The amino acid sequence of the *LIP1* product is shown in Fig.1: it is consistent with a protein of 549 residues, the first 15 of them building a canonical leader sequence for secretion, a stretch of hydrophobic amino acids that targets the growing peptide to the secretory pathway. Sequencing of the mature protein N-terminus showed that this signal peptide is cut in *Candida* after the tripeptide VAA [8]. The mature protein contains 3 potential sites for N-glycosylation in the correspondence of the N-X-T/Y consensus sequence at asparagine 291, 314 and 351.

```
                    ▼
1   MELALALSLI ASVAAAPTAT LANGDTITGL NAIINEAFLG IPFAEPPVGN
51  LRFKDPVPYS GSLDGQKFTS YGPSCMQQNP EGTYEENLPK AALDLVMQSK
101 VFEAVSPSSE DCLTINVVRP PGTKAGANLP VMLWIFGGGF EVGGTSTFPP
151 AQMITKSIAM GKPIIHVSVN YRVSSWGFLA GDEIKAEGSA NAGLKDQRLG
201 MQWVADNIAA FGGDPTKVTI FGESAGSMSV MCHILWNDGD NTYKGKPLFR
251 AGIMQSGAMV PSDAVDGIYG NEIFDLLASN AGCGSASDKL ACLRGVSSDT
301 LEDATNNTPG FLAYSSLRLS YLPRPDGVNI TDDMYALVRE GKYANIPVII
351 GDQNDEGTFF GTSSLNVTTD AQAREYFKQS FVHASDAEID TLMTAYPGDI
401 TQGSPFDTGI LNALTPQFKR ISAVLGDLGF TLARRYFLNH YTGGTKYSFL
451 SKQLSGLPVL GTFHSNDIVF QDYLLGSGSL IYNNAFIAFA TDLDPNTAGL
501 LVKWPEYTSS SQSGNNLMMI NALGLYTGKD NFRTAGYDAL FSNPPSFFV
```

Figure 1. Amino acid sequence of lipase 1. The arrowhead marks the site for signal peptidase cleavage. Glycosylation sites are underlined.

The first step towards heterologous expression was the modification by polymerase chain reaction of the 5'-end of *LIP1* in order to add an ATG start codon, since the original triplet had been removed during cloning [1]. The complete *LIP1* open reading frame was inserted in pEMBLYex4. In the resulting plasmid *LIP1* is put under the control of the UASGal regulation sequence: its expression is modulated through the carbon source present in the culture medium, i.e. repressed by glucose and induced by galactose.

S. cerevisiae cells transformed with the recombinant plasmid were induced by addition of galactose and analysed for both the production of specific mRNA and of recombinant protein. Surprisingly, and despite the presence of abundant mRNA as revealed by Northern blot analysis, no trace of the recombinant protein could be found neither in the cells nor in the culture medium (Fig. 2).

A different expression vector was obtained by replacing the original signal peptide (amino acids 1-15 of LIP1) with the leader sequence derived from the killer toxin (KI) from another yeast, *Kluyveromyces lactis*. This sequence proved to be very efficient in the expression/secretion of foreign protein such as human interleukin 1, as described in [9].

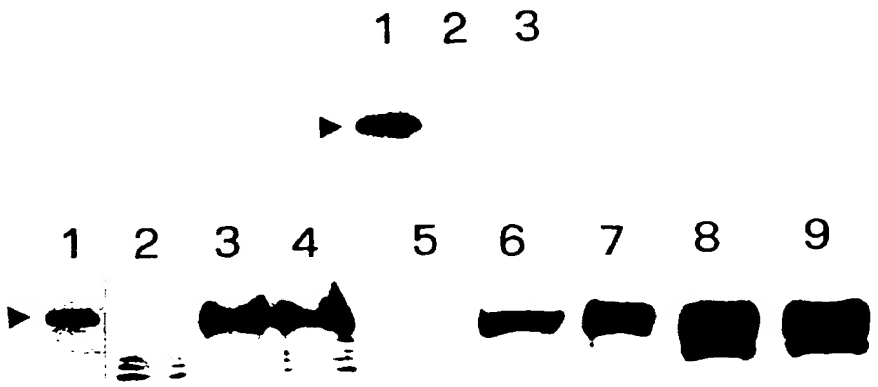

Figure 2. Western blotting of cell extracts of *S. cerevisiae* cells transformed with inducible plasmids containing *LIP1* preceeded either by its leader sequence (top) or by that derived from the *K. lactis* killer toxin (bottom). **Top** (homologous leader sequence) cell extracts from cells grown either in glucose (2) or in galactose minimal medium (3). Lane 1 contain 1 μg of purified lipase. **Bottom** (heterologous l.s.): cells grown either in glucose (2) or in galactose medium to the exponential or stationary phase of growth (3 and 4, respectively). (5-9) Production of recombinant lipase in bioreactor under repression (5) or induction conditions (6-9). Lane 1 contain 1 μg of purified lipase. Extracts were obtained from 10^6 cells.

When yeast cells transformed with the chimeric plasmid were induced by galactose, high levels of specific mRNA were detected by Northern blot analysis and recombinant lipase I was accumulated within the host cells (Fig.2). The enzyme was synthesized over several hours during the exponential and the stationary phases of growth. Its level was evaluated to be about 10-20 mg/liter of culture, a production comparable to that obtained in our laboratory with similar expression systems using batch fermentation.

The process was scaled up in a two-stage computer-controlled fermentation system [10]. Fed batch fermentation was performed in a 2 l aereated, stirred-tank bioreactor equipped with temperature, agitation, air flow rate, dissolved oxygen, pH and ethanol controllers. The control system was based on the determination of the ethanol concentration in the bioreactor outflow gases, allowing to regulate the supply rate of nutrients to the culture so as to prevent inhibition of cell growth. In this way, a high yield of conversion of glucose to biomass (0.48 g/g) and an elevated volumetric production and productivity can be obtained (up to 100 g of dry weigth per liter) [10].

To induce *LIP1* expression, a mixed substrate was added in the feed: the process was started with 50% w/v glucose; induction was obtained by changing the carbon source to 25%glucose-25%galactose after 40 hours and to 50% galactose after 80 hours. Fermentation was stopped after 115 hours when cell density in the bioreactor was 10^9 cells/ml. Up to 1 g/l of recombinant lipase I was obtained and the protein was accumulated over several hours of fermentation (Fig. 2).

In conclusion, these experiments suggest that the lipase natural leader sequence, although very efficient in *Candida* cells, is not suited for expression in *S. cerevisiae*. *S. cerevisiae* cells are usually quite permissive in the recognition of foreign signals and several eukaryotic proteins preceeded by their own signals have been successfully expressed and secreted . However, a correct processing of the recombinant protein cannot be assumed *a priori*, since for example, expression of interferon preceeded by its natural leader sequence resulted in protein degradation [11]. Whereas the role of the leader sequence in secretion is well established since several years, its effect on protein translability/stability is still uncertain, also because it seems to depend on the protein itself.

To investigate the fate of the recombinant protein within the yeast cells, cell extracts were subjected to treatment with endoglycosidase F to remove carbohydrates. A decrease in the molecular weight of the recombinant protein was evidenced in Western blots (Fig.3), showing that lipase had been correctly targeted to the secretory pathway, since glycosylation first occurs in the endoplasmic reticulum. The recombinant lipase however, was retained in the cell membrane and was not released neither in the culture medium nor in the periplasmic space (Fig. 3). This is very likely to depend on a misfolding of the recombinant protein due to the erroneous introduction of leucine residues in the

correspondence of CUG codons, a problem deriving from the non universal usage of this codon by *C. rugosa* cells, as discussed in the following.

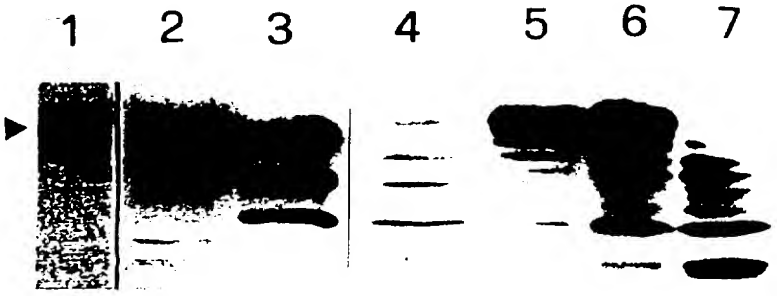

Figure 3. Transport of rLIP1 in yeast cells. Extracts were obtained from 10^7 cells exponentially growing in galactose (induction) medium and subjected to Western blotting. (2) untreated sample; (3) after deglycosylation with endoglycosidase F; (4-7) separation of cells fraction: membrane fraction of cells grown in glucose (4) or galactose (5-6) and cytoplasmatic fraction of cells induced by galactose. The control purified protein is shown in 1.

2. Mutagenesis of CUG codons

In 1989 Kawaguchi and colleagues unexpectedly reported that codon CUG was read as serine instead of leucine by *C. rugosa*. This finding was obtained by comparing the content of serine and leucine predicted from the nucleotide sequence of *LIP1* cDNA and that experimentally determined [8]. This was the first example of a deviation from the universal genetic code in nuclear genes, since previously this phenomenon was believed to be restricted to mithocondrial genomes [12]. Shortly later, CUG-Ser was found also in other *Candida* species, namely *C. albicans*, *C. parapsilosis*, *C. melibiosica*, *C. zeylanoides* and *C. maltosa* [13,14].

This might seriously hinder the expression of *Candida* genes in any host organism, since the translational machinery of the host cells would introduce leucine residues at any CUG codon, where a serine was present in the original protein. This proved to be not crucial for several *Candida* proteins that were cloned by complementation of *Saccharomyces* mutations, as for example the orotidine 5'- phosphate decarboxylase from *C. maltosa* [15]. All these genes carried only a few CUG codons, apparently at positions where a substitution can be accepted.

In *C. rugosa* lipases however, CUG codons are very frequent and even the catalytic serine is CUG-encoded (Fig.4). Thus, an extensive mutagenesis program might be necessary to obtain the expression of a functional enzyme. In the following are presented preliminary results of this study that was performed with two main purposes: the production of a functional recombinant lipase and the investigation of basic issues on the molecular evolution of CRL and the origin of the deviation from the universal genetic code.

CRL genes appear to be unique within the *Candida* genus for their high usage of the codon CUG (see table I) that is employed for 40% of all serine codons, in contrast to several other *Candida* species, where CUG account for only 2-3% of total serine codons (G. Pesole et al., unpublished results).

Table I - Frequencies of codons for serine and leucine in *C. rugosa* lipases(1)

		LIP1	LIP2	LIP3	LIP4	LIP5
Serine						
	CUG	0.42	0.38	0.41	0.36	0.37
	AGU	0.02	0.09	0.09	0.04	0.06
	AGC	0.25	0.29	0.28	0.29	0.28
	UCG	0.17	0.13	0.13	0.16	0.11
	UCA	0.02	0.02	0.00	0.02	0.00
	UCU	0.00	0.02	0.02	0.02	0.02
	UCC	0.11	0.07	0.06	0.08	0.15
Leucine						
	UUG	0.44	0.43	0.47	0.44	0.42
	UUA	0.00	0.00	0.00	0.00	0.00
	CUA	0.00	0.00	0.00	0.00	0.00
	CUU	0.08	0.21	0.10	0.15	0.15
	CUC	0.48	0.36	0.43	0.40	0.43

(1) Values are given as percentages of total codons for each amino acid

In the LIP1 protein, 19 serines (enclosed the catalytic serine 209) are CUG-serines. By comparing the amino acid sequences of the five CRL enzymes sequenced to date, one can observe that 13 CUG serines are conserved in all isoforms (Fig.4). This might be suggestive of their importance in CRL structure. A broader analysis carried out with enzymes belonging to the esterase/lipase family which includes the *C. rugosa* and *G. candidum* lipases together with esterases and acetylcholinesterases [16], shows that at least three of these serines: S209, S241 and S282 are conserved also in closely related enzymes. The structural role of these residues is confirmed by the inspection of their position in the enzyme three-dimensional structure. Serine 241 for example, interacts with two catalytic residues, S209 and E341. These codons were therefore mutated by polymerase chain reaction to universal serine TCC triplets and reintroduced in the *LIP1* gene cloned in the expression plasmid described above. A high level of expression was obtained in experimental conditions very similar to those described in the previous paragraph but the protein was accumulated within the cells in an inactive form.

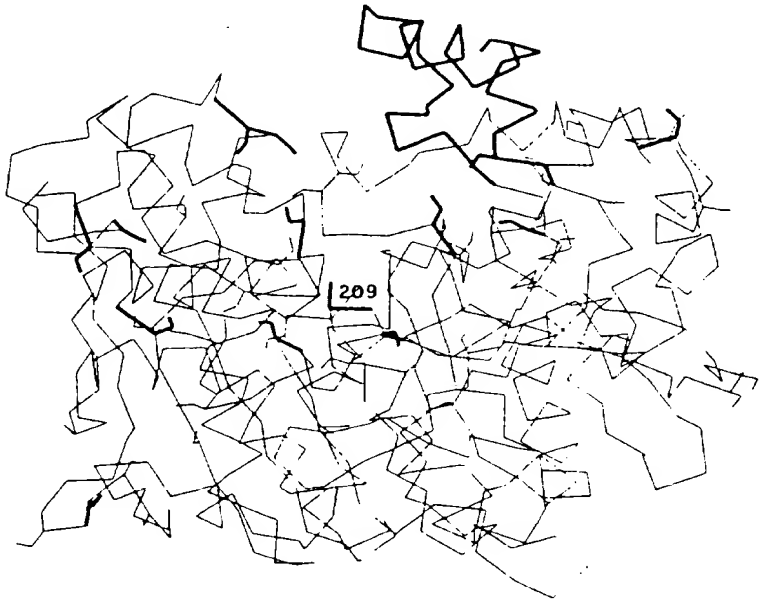

Figure 4. CA trace of CRL1 with the conserved serines and the lid in bold. The figure was generated by Turbo-Frodo [20]

A second group of target residues was identified taking into account their structural role according to the hydrogen bonds and salt bridges pattern and mutagenized. Once more, the protein was retained in the endoplasmic reticulum, although serines important for the structure of the protein core had been restaured

These results seem to suggest that recovery of the inner core conformation is not sufficient to obtain a functional enzyme.

Some intriguing questions arise from the study of *C. rugosa* lipases, in particular how could such a deviation from the universal genetic code originate and be maintained in the lipase enzymes where it concerns even the catalytic residue.

By aligning the sequences of the small ribosomal subunits rRNAs of all known *Candida* species and other related yeast [17], we could observe that the genus *Candida* includes a monophyletic subgroup, several members of which have been shown by Ohama and colleagues [13] to use CUG as a codon for serine. Interestingly, other members of this group only rarely employ CUG which is the major serine codon in *C. rugosa*. This finding can be partly explained by the high C+G content of the *C. rugosa* genome and by the existence of multiple genes for tRNA(CAG)[18], but does not satisfactory explain how a non-functional mutation could arise and be maintained in an enzyme active site. In fact, we should hypothesize that a non-functional lipase originated following CUG unassignment. One can hypothesize that either codon change was made possible by the presence of multiple copies of lipase genes so that a deleterious mutation would not affect the viability of the cell, or that lipase was a dispensable function. We have shown that *C. rugosa* possess multiple copies of the lipase genes and it is a diploid organism [19]. Thus it was possible for non functional copies of this gene to be maintained within the genome and to assume their functional role after codon reassignment.

Both the high number of CUG serines and their conservation in the CRL gene family suggest the action of a strong selective pressure on the assignment/usage of this codon.

Acknowledgements

This work was supported by the Progetto Finalizzato Biotecnologie e Biostrumentazioni of the Italian National Research Council and by the BRIDGE Program of the Commission of the European Communities.

was generated by Turbo-Frodo [20].

3. References

1. Longhi, S., Fusetti, F., Grandori, R., Lotti, M., Vanoni, M. and Alberghina, L. (1992) Cloning and nucleotide sequences of two lipase genes from *Candida cylindracea*, *Biochim. Biophys. Acta* **1131**, 227-232.
2. Lotti, M., Grandori, R., Fusetti, F., Longhi, S., Brocca, S., Tramontano, A. and Alberghina, L. (1993) Cloning and analysis of *C. cylindracea* lipase sequences, *Gene* **124**, 45-55.
3. Rua, M.L., Diaz-Maurino, T., Fernandez, V.M., Otero, C. and Ballesteros, A. (1993) Purification and characterization of two distinct lipases from *Candida cylindracea*, *Biochim. Biophys. Acta* **1156**, 181-189.
4. Grochulski et al., (1993) Insights into interfacial activation from an open structure of *Candida rugosa* lipase, *J. Biol. Chem.* **268**: 12843-12847.
5. Alberghina, L., Lotti, M., Martegani, E., Ranzi, B.M. and Porro, D. (1993) Heterologous gene expression in budding yeast: from micrograms to grams/litre, *Med. Fac. Landbouww. Univ. Gent* **58/4b**, 1901-1909.
6. Cesareni, G. and Murray, J.A.H. (1987) in: *Genetic Engineering Principles and Practice* Vol 5, Plenum Press, New York.
7. Johnston, M. (1989) A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*, *Microbiol. Review* **51**, 458-496.
8. Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. and Iwasaki, S. (1989) The codon CUG is read as serine in an asporogenic yeast *Candida cylindracea*, *Nature* **341**, 164-166.
9. Baldari, C., Murray, J.A.H., Ghiara, P., Cesareni, G. and Galeotti, V. (1987) A novel leader peptide which allows efficient secretion of a fragment of human interleukin 1β in *S. cerevisiae*, *EMBO J.* **6**, 229-234.
10. Alberghina, L., Porro, D., Martegani, E., and Ranzi, B.M. (1991) *Biotechnol. Appl. Biochem.* **14**, 82-89.
11. Romanos, M.A., Scorer, C.A. and Clare, J.J. (1992) Foreign gene expression in yeast: a review. *Yeast* **8**, 423-488.
12. Barrel, B.G., Bankier, A.T. and Drouin J. (1979) A different genetic code in human mitochondria, *Nature* **282**, 189-194.
13. Ohama, T., Suzuki, T., Mori, M., Osawa, S., Ueda, T., Watanabe, K. and Nakase, T. (1993). Non-universal decoding of the leucine codon CUG in several *Candida* species, *Nucl. Acid Res.* **21**, 4039-4045.
14. Sugiyama, H., Ohkuma, M., Masuda, Y., Park, S.-M., Ohta, A. and Takagi, M. (1995) *In vivo* evidence for the non-universal usage of the codon CTG in *Candida maltosa*, *Yeast* **11**, 43-52.
15. Ohkuma, M., Muraoka, S., Hwang, C.W., Ohta, A. and Takagi, M. (1993) Cloning of the C-URA3 gene and construction of a triple auxotroph (*his5*, *ade1*, *ura3*) as a useful host for the genetic engineering of *Candida maltosa*, *Curr. Genet.* **23**, 205-210.
16. Cygler, M., Schrag, J.D., Sussman, J.L., Harel, M., Silman, I., Gentry, M.K. and Doctor, B.P. (1993) Relationship between sequences conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins, *Prot. Sci.* **2**, 366-382.

17. Alberghina, L., Lotti, M., Brocca, S., Pesole, G. and Saccone, C. (1995) Protein Engineering of *Candida rugosa* lipase: replacement of CUG serine residues and folding. Proceedings of the 7th European Congress on Biotechnology, Elsevier, in press.
18. Suzuki, T., Ueda, T., Yokogawa, T., Nishikawa, K. and Watanabe, K. (1994) Characterization of serine and leucine tRNA in an asporogenic yeast *Candida cylindracea* and evolutionary implications of genes for $tRNA^{SerCAG}$ responsible for translation of a non-universal genetic code, *Nucl. Acid Research*, **22**, 115-123.
19. Brocca, S., Grandori, R., Breviario, D. and Lotti, M. (1995) Localization of lipase genes on *Candida rugosa* chromosomes, Curr. Genetics, in press
20. Turbo-Frodo version 4.4 by A. Roussel and A.G. Inisan, Bio-Graphics, Marseille, France.